

Codarea imaginilor bazată pe regiuni, pentru aplicații multimedia

Conf. dr. Florin Alexa
Universitatea Politehnica din Timișoara

Abstract. Region based video coding has been a very active research area over the last few years. It has been viewed as a potential alternative to traditional schemes suffering from the “blockiness” at very low bit rates. The success of region based techniques relies on the ability to describe regions in an image succinctly by their shape and sizes. In this paper we present a new approach to region based image coding. It is based by incorporation the knowledge of the border in image segmentation and exploits temporal redundancies between successive images.

Keywords: image processing, segmentation algorithm, region based image coding.

1 Introducere

Ultimii ani au constituit o etapă cu o dezvoltare spectaculoasă a domeniului comunicațiilor. Au fost introduse foarte multe servicii telematice și de televiziune ce au dus la creșterea numărului de imagini transmise la diferite niveluri. Progresele înregistrate în electronică, în tehnologiile software și, nu în ultimul rând, corelarea eforturilor de cercetare ale specialiștilor din domenii distincte ca electronica, grafica computerizată, vederea artificială etc., au făcut ca azi să se vorbească tot mai mult despre: supermagistrale de informații, aplicații multimedia supersofisticate, echipamente performante de comunicații interpersonale și multe altele. Traficul de informație pe rețelele de calculatoare existente prin intermediul serviciilor Internet ca: E-Mail, Telnet, FTP, Usenet, WWW, Gopher etc., a crescut continuu putând spune că se marchează startul spre o nouă eră: *era informației*. Toate acestea demonstrează clar că industria de telecomunicații capătă o nouă formă și ocupă o poziție tot mai importantă în cadrul societății moderne.

În acest context compresia digitală a imaginilor, utilizând tehnici performante, a cunoscut o dezvoltare impresionantă în decursul ultimilor

decenii. S-au elaborat și propus foarte multe standarde atât pentru secvențele de imagini: H-261, H-263, MPEG-1, MPEG-2, MPEG-4, MPEG-7, cât și pentru imagini statice: JPEG, JPEG-2000, multe dintre ele constituind elementul cheie al noilor tehnologii informatice în diverse aplicații: multimedia, efecte speciale în televiziune și cinematografie, grafică pe calculator, sisteme de videoconferință, teledetecție, telemedicină etc. Cele mai multe dintre aceste servicii necesită transmiterea unei cantități impresionante de date pe canale foarte înguste.

Codarea la rate foarte scăzute de biți, necesară transmisiei pe canalele de comunicații existente, implică rate de transmisie sub 64 kb/s pentru semnalul video. Sursele video aferente acestor aplicații, având 8 biți/eșantion pentru fiecare culoare, conțin în mod obișnuit aproximativ 1,25 Mocteți pentru un cadru de imagine color. Cum frecvența cadrelor este de 25 Hz (30 Hz), rezultă o rată de date, fără implicarea unei compresii, de peste 10 Mb/s. În aceste condiții compresia semnalelor trebuie să fie cuprinsă între 150:1 și 1000:1.

Cu toate că se obțin performanțe bune din punct de vedere al calității obiective (eroare medie pătratică, raport semnal/zgomot), tehnicile clasice de corespondență a blocurilor [MPG85], [DM95], [XPC99] suferă din cauză că, datorită reprezentării sub formă de pixeli sau ca blocuri de pixeli, aceste imagini nu au nici o semnificație din punct de vedere semantic. Astfel, la rate de compresie ridicate se obțin erori majore din punct de vedere al sistemului vizual uman. Din acest motiv nu pot fi obținute îmbunătățiri majore ale performanțelor compresiei.

În plus, lipsa informațiilor referitoare la conținutul imaginii limitează funcționalitățile disponibile la recepție, nefiind posibilă manipularea și personalizarea informației recepționate, așa cum se dorește în cadrul sistemelor de transmisie multimedia.

În contrast cu tehnicile clasice de codare tehnicile bazate pe reprezentarea semantică [AH95], [KC98], au încercat imitarea funcțiilor sistemului vizual uman. În acest scop algoritmi se bazează pe descrierea imaginii sub forma obiectelor care o compun. Compresia se obține luând în considerare proprietățile sistemului vizual uman și anume alocarea mai multor biți în zonele importante și reducerea numărului acestora în rest.

Ca rezultat al folosirii reprezentării semantice a imaginii, aceste tehnici permit o rată de compresie foarte ridicată, cu obținerea unor deteriorări vizuale ale imaginii foarte mici, iar informațiile despre conținutul imaginii oferă funcționalități suplimentare la recepție. În particular, utilizatorului final îi este accesibilă interactivitatea.

În aceste condiții robustețea estimării mișcării este mult mai importantă decât obținerea unei predicții temporale ușoare, în vederea asigurării unei calități corespunzătoare a secvenței de imagini reconstituite.

Pentru obținerea unei rate de compresie ridicată este necesară determinarea unui câmp al mișcării reprezentativ deoarece:

- se dorește eliminarea erorilor de predicție care sunt transmise către decodor, motiv pentru care este important să nu se introducă distorsiuni suplimentare deranjante pentru sistemul vizual uman,
- se dorește subeșantionarea temporală a secvenței de imagini înaintea codării.

În cele ce urmează este prezentat un algoritm de codare bazat pe estimarea vectorilor de mișcare în vederea exploatarei redundanței temporale a secvențelor de imagini.

2 Codarea imaginilor

Principalul argument în favoarea tehnicilor bazate pe corespondența blocurilor îl reprezintă schemele de codare hibridă, care combină predicția cu compensarea mișcării în domeniul timp, cu tehnicile de decorelație în domeniul spațial. Aceste scheme sunt utilizate în mod curent în standardele actuale (H261, H263, MPEG-1, MPEG-2) și pot fi utilizate ca puncte de start pentru tehnicile de codare bazate pe regiuni.

Principalele dezavantaje ale acestor scheme de codare sunt:

- divizarea “oarbă” fără a lua în considerație conținutul semantic al imaginii, care determină apariția efectelor de bloc atunci când sunt cerute compresii mari;
- modelele de mișcare sunt aplicate unor blocuri ce nu corespund în totalitate cu mișcarea obiectelor constituente ale imaginii;
- este dificilă încorporarea proprietăților sistemului vizual uman în aceste scheme de codare.

O posibilitate de eliminare a acestor dezavantaje este încorporarea proprietăților sistemului vizual uman în algoritmi de codare, și anume:

- informația referitoare la contur și muchii este apreciată foarte mult de sistemul vizual uman, fiind responsabilă principală în ceea ce privește percepția;
- textura are o importanță diminuată, fiind asociată cu informația suplimentară, influențând percepția numai atunci când este asociată cu informația de contur.

- cele mai multe dintre obiectele naturale pot fi approximate ca fiind membri ai unor grupări (clase) deterministice, pe baza unor seturi de caracteristici.

Problemele legate de: segmentarea spațio-temporală și estimarea mișcării bazată pe segmentare și urmărire au făcut obiectul cercetărilor în domeniile analizei secvențelor de imagini și al vederii artificiale. De altfel, codarea secvențelor de imagini implică abordări foarte diferite deoarece scopul este acela de a realiza o reprezentare cât mai compactă a informației de mișcare, motiv pentru care este foarte importantă precizia de estimare a mișcării și cantitatea de informații suplimentare ce vor fi transmise spre decodor [Gir94], [MDN93]. Aceasta implică următoarele:

- controlul procesului de segmentare;
- posibilitatea alegerii modelului de mișcare;
- cuantizarea parametrilor de mișcare.

De asemenea, în scopul utilizării secvențelor codate în cadrul serviciilor interactive, este necesară considerarea întârzierii care se obține în procesul de codare.

Cea mai simplă implementare a codării bazate pe regiuni o reprezintă codarea erorii, care utilizează redundanța temporală existentă în secvențele de imagini. Astfel, fiecare cadru este obținut prin predicția celui anterior, semnalul ce este codat fiind eroarea (diferența) dintre cele două cadre: cel curent și cel predictat (figura 2.1).

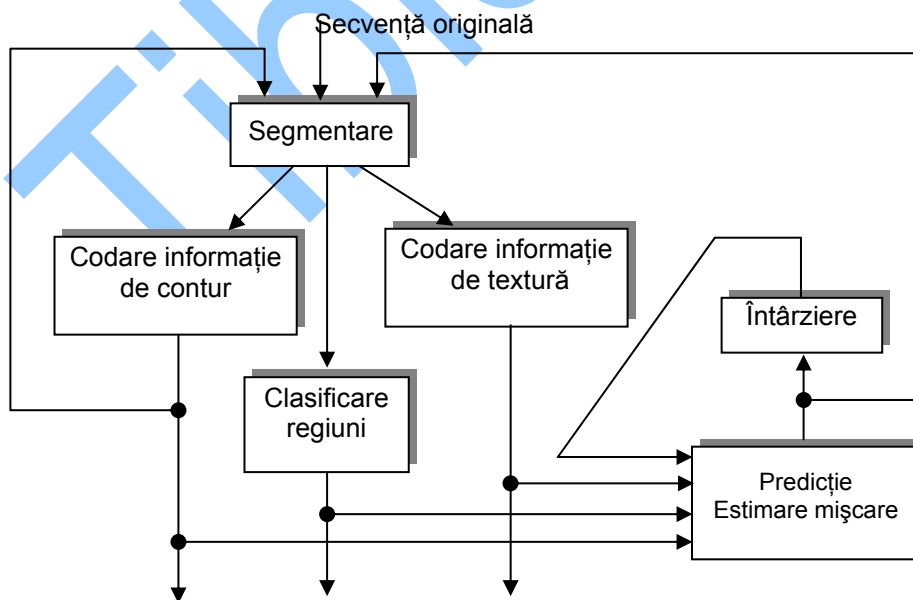


Fig.2.1 Schema bloc a codorului bazat pe regiuni

Predicția este una simplă, iar pentru fiecare parametru ce se transmite (contur, clasificare, textură) se folosesc proceduri de codare specifice. Pentru formă se utilizează tehnici specifice codării conturilor, cu excepția modificărilor temporale, caz în care se utilizează codarea predictivă. Pentru mișcare este utilizată codarea predictivă, în raport cu parametrii de mișcare estimați pentru același obiect în imaginea anterioară. Pentru informația de textură se utilizează tehnici bazate pe forma de undă (codare hibridă transformare-predicție, cu compensarea mișcării).

3 Algoritm de codare bazat pe regiuni

Având în vedere tendința actuală spre codarea la rate foarte scăzute de biți cu aplicații în video-telefonie, video-conferință, multimedia, este clar că soluția o reprezintă implementarea tehnicilor de codare bazate pe regiuni [CP98], [MCR02].

În acest scop autorul a propus un algoritm de estimare a mișcării pentru codarea secvențelor de imagini care se bazează pe o segmentare statică adaptivă a imaginilor [Ale98]. Schema bloc a algoritmului propus pentru estimarea și codarea secvențelor de imagini este prezentată în figura 3.1

Ca date de intrare algoritmul utilizează imaginea curentă originală și segmentată. De asemenea, există o întârziere de două cadre înaintea generării primei imagini codate, întrucât se utilizează o codare cu predicție, motiv pentru care este necesară cunoașterea atât a imaginii transmise cât și a imaginii curente.

Problema principală care se pune este modul în care poate fi descris un obiect care se mișcă în secvență în așa fel încât să se obțină o compresie cât mai mare a informației. Tehnica propusă este o tehnică de căutare globală pe baza unor parametri de formă și de poziție. Pe baza imaginilor originale curente și precedente, a imaginilor segmentate și a caracteristicilor regiunilor se realizează estimarea mișcării prin urmărirea regiunilor și determinarea corespondențelor dintre acestea.

Pe baza vectorilor de mișcare se realizează o compresie cu predicție a informației, transmițându-se eroarea, vectorii de mișcare, eticheta regiunii și corespondența precum și coeficienții funcției de aproximare a luminanței. În cazul în care regiunile nu au corespondență în cadrul anterior se transmite conturul regiunii precum și informația referitoare la coeficienții funcției de aproximare a luminanței. Algoritm de are posibilitatea unei codări ierarhice în sensul că se poate transmite și informația referitoare la textură dacă se dorește o codare fără pierderi.

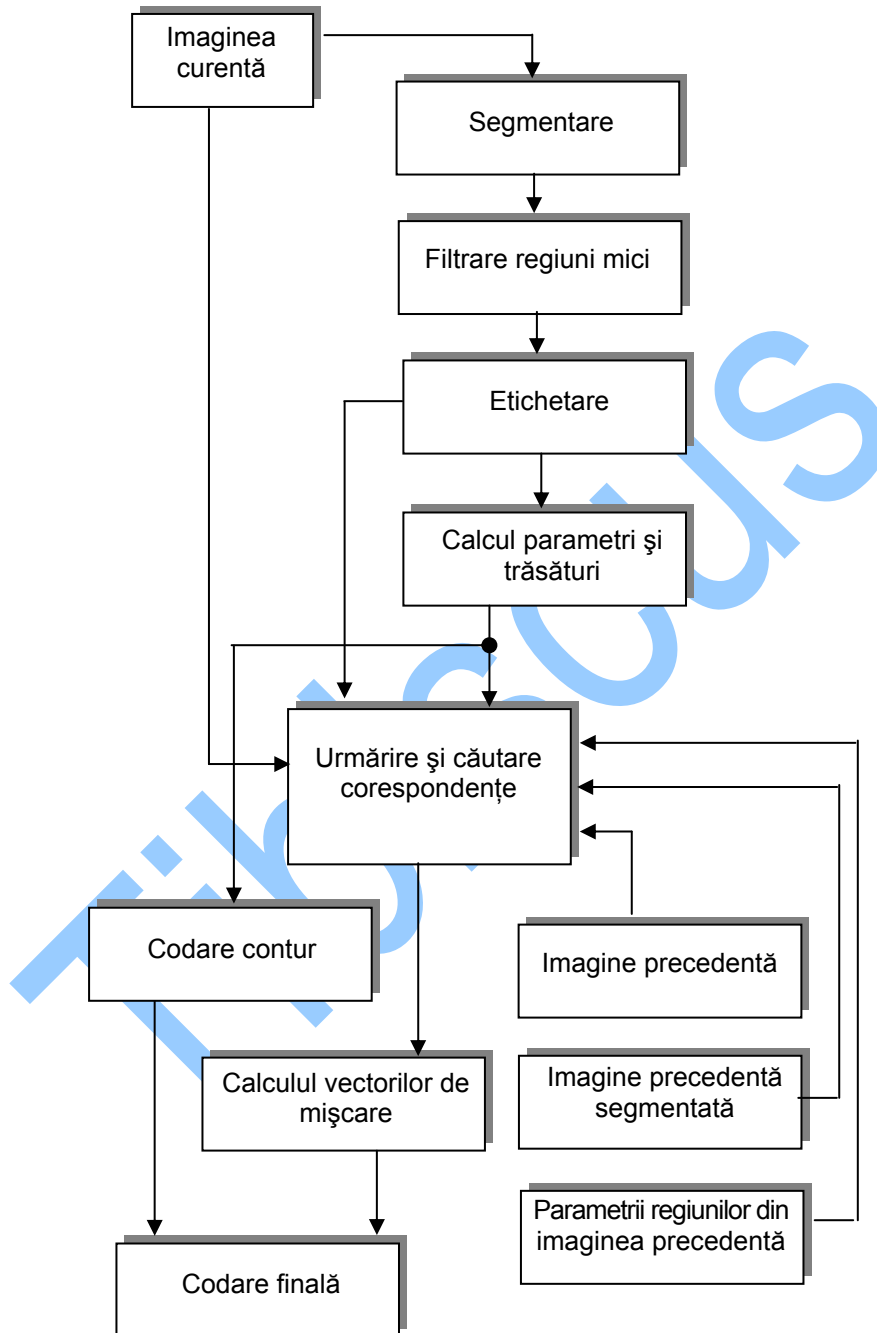


Figura 3.1 Algoritm de estimare a mișcării în vederea codării secvențelor de imagini la rate foarte scăzute

Modelul de descriere a mișcării care se alege în cazul acestei aplicații este un model parametric definit prin comparația parametrilor de descriere a regiunilor (poziția centrelor de greutate pentru translație, unghiul dintre axa de alungire și ordonată sau unghiul de rotație determinat prin metoda cercului pentru rotație, raportul caracteristicilor de formă pentru scalare).

Pentru modelul de mișcare ales se impun o serie de limitări. Acestea sunt: valorile maxime ale translației (16 pixeli), rotației (20 grade) și măririi/micșorării (50%), și se alege pe baza considerentului vitezei maxime de deplasare a obiectelor și modificării maxime a distanței focale pentru obținerea a două imagini succesive.

Aceste limitări sunt necesare pentru a se evita corespondența între regiuni asemănătoare din punctul de vedere al formei, dar care nu corespund din punctul de vedere al mișcării reale. Apariția de corespondențe false, poate fi explicată prin faptul că pot apărea regiuni noi în imagine, care sunt asemănătoare cu regiuni din imaginile anterioare.

Varianta de căutare a corespondențelor aleasă pentru acest algoritm este una globală în cadrul imaginilor. Ca date de intrare algoritmul de căutare utilizează imaginea precedentă și cea curentă precum și structurile cu parametri de descriere a regiunilor din imagini. Se parcurg regiunile din imaginea curentă. Pe baza parametrilor de poziționare a regiunilor în imagine și pe baza parametrului de formă definit anterior se determină regiunile din imaginea precedentă care asigură cea mai bună corespondență cu regiunea curentă. Decizia finală de corespondență se ia pe baza calculului erorii medii pătratice.

Algoritmul de căutare este prezentat în figura 3.2 și se realizează pentru toate regiunile din imaginea curentă prin parcurgerea lor o singură dată. În scopul reducerii timpului necesar pentru calcule, regiunile pentru care a fost stabilită corespondența nu mai sunt luate în calcul pentru corespondențele regiunilor următoare.

Datele obținute la căutarea corespondențelor sunt utilizate pentru determinarea vectorilor de mișcare. Parametrii determinați sunt:

- translația, prin diferența dintre coordonatele centrelor de greutate;
- rotația, prin estimarea unghiului de rotație;
- scalarea (mărirea/ micșorarea), prin determinarea raportului dintre ariile celor două regiuni.

Acest algoritm realizează o compresie prin: transmiterea vectorilor de mișcare pentru imaginea curentă, în cazul regiunilor având o corespondență în imaginea precedentă și transmiterea conturului, centrului de greutate și modelului texturii pentru regiunile ce nu au o corespondență în imaginea precedentă.

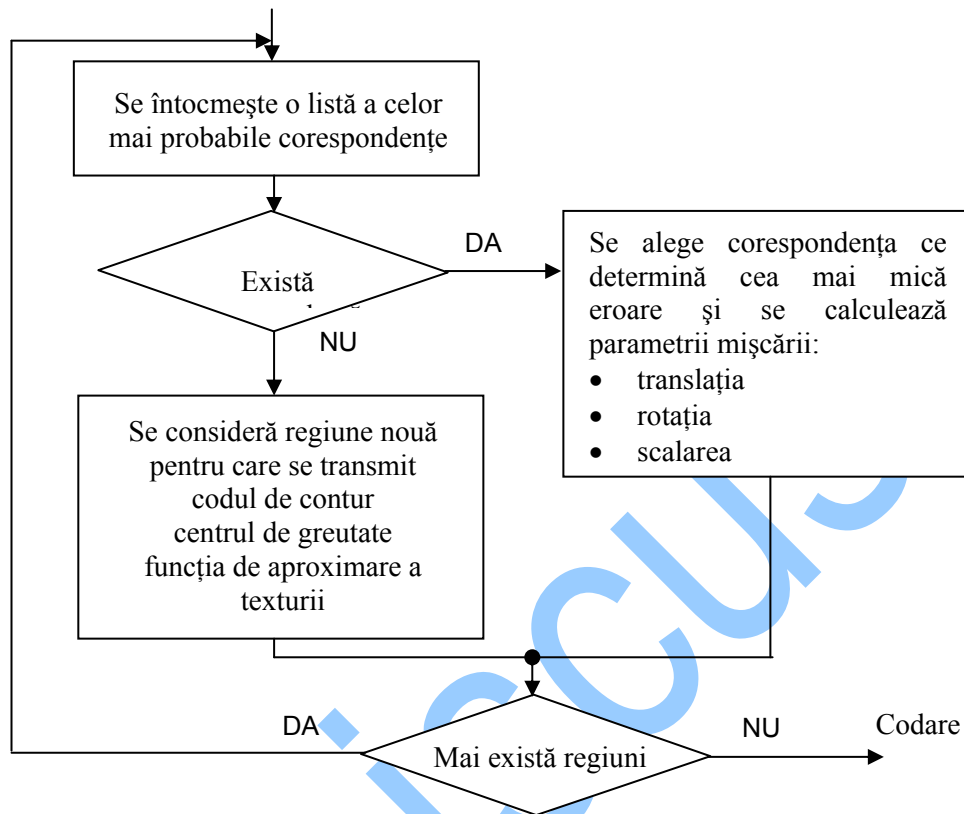


Figura 3.2 Organigrama algoritmului de corespondență al regiunilor

Algoritmul de refacere a imaginii se bazează pe refacerea unei imagini aparținând unei secvențe pe baza parametrilor transmiși și ai imaginii precedente. Pentru a se putea reface imaginea se consideră că există vectorii de mișcare, imaginea precedentă și imaginea sa de etichete din parcurgerea anterioară a algoritmului de refacere.

4 Rezultate experimentale

Algoritmul propus pentru estimarea mișcării se bazează pe utilizarea reprezentării semantice a imaginii. Pentru determinarea regiunilor (obiectelor) din imagine se poate utiliza o segmentare: spațială, temporală sau spatio-temporală a imaginilor. Imaginile utilizate pentru evaluarea algoritmului sunt imagini standard: “tenis”, “claire”, “salesman” și “missa”. Figurile 4.1 și 4.2 prezintă raportul semnal/zgomot și factorii de compresie obținuți pentru diferitele secvențe analizate.

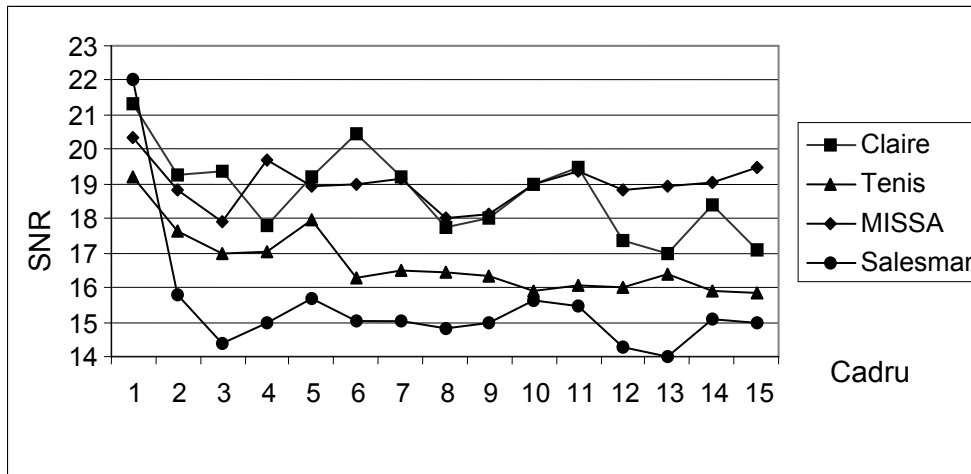


Figura 4.1 Raportul semnal/zgomot obținut pentru imaginile reconstruite pentru diferite secvențe

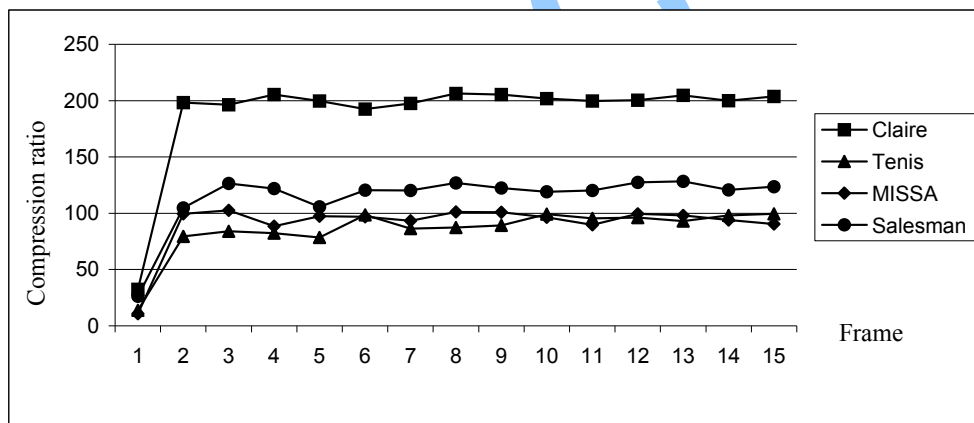


Figura 4.2 Factorii de compresie obținuți pentru diferitele secvențe

Se observă că imaginile reconstruite figurile 4.3.b și 4.4.b, din compararea cu imaginile originale, permit distingerea trăsăturilor esențiale ale personajelor prezente în imagini. Ținând cont de factorul de compresie relativ mare obținut, 192,17:1, putem considera că algoritmul propus satisface necesitățile pentru o transmisie la rate foarte scăzute de biți.

Pentru imaginea Claire reprezentată în formatul QCIF (176×144), rata de transmisie obținută este 0,0408 biți/pixel, ceea ce este echivalent cu o rată de transmisie de 6,204 kbiți/s la o rată de 6 cadre/s. Rezultatele obținute ne îndreptătesc să considerăm că se pot obține rezultate acceptabile

chiar și la transmiterea imaginilor color. Pentru celelalte imagini, factorul de compresie mediu obținut pentru primele 15 cadre din secvența este:

- Tennis: 85,39
- MISSA: 90,56
- Salesman: 114,29.

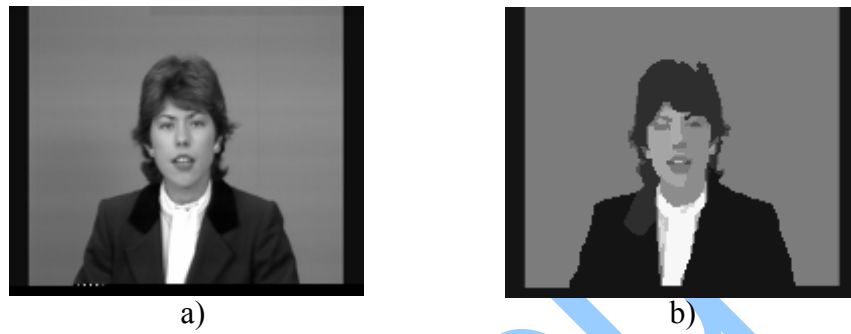


Fig.4.3. Cadrul 5 original și respectiv reconstituit din secvența „Claire”

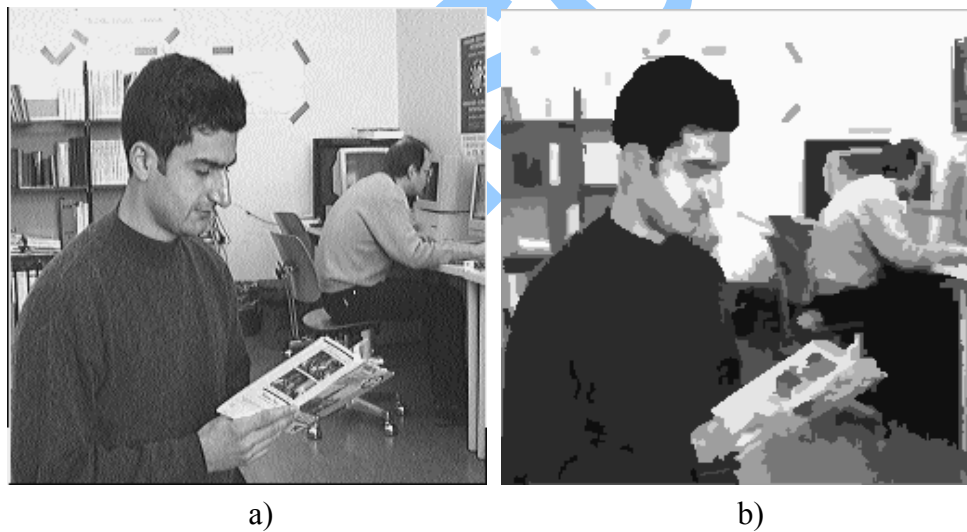


Fig.4.3. Cadrul 15 original și respectiv reconstituit din secvența „Missa”

5 Concluzii

În concluzie, se poate spune că metodele de estimare a mișcării bazate pe regiuni, cu încorporarea proprietăților sistemului vizual uman sunt potrivite

pentru tehnicile de codare la rate foarte scăzute, care necesită rate de compresie foarte mari.

Algoritmul propus realizează estimarea mișcării bazându-se pe proprietățile sistemului vizual uman:

- informația referitoare la contur și muchii este apreciată foarte mult de sistemul vizual uman, fiind responsabilă principală în ceea ce privește percepția;
- textura are o importanță diminuată, fiind asociată cu informația suplimentară, influențând percepția numai atunci când este asociată cu informația de contur, putând fi transmisă, de cele mai multe ori, doar cu ajutorul unui bit.

Algoritmul de segmentare conduce la partiționarea imaginii în regiuni de formă oarecare, cu ajutorul cărora, pe baza corespondențelor dintre parametri caracteristici, se estimează vectorii de mișcare.

Atribuirea adaptivă a modelului de mișcare pentru fiecare regiune se bazează pe presupunerea că scena cuprinde numai obiecte rigide (corpuri nedeformabile) și există limite maxime ale mișcării între imaginile succesive. De asemenea este utilizat un set de trăsături caracteristice cu ajutorul cărora este posibilă determinarea translației, scalării și rotației regiunilor.

Rezultatele obținute cu o complexitate moderată, factori de compresie de aproximativ 130:1 și raport semnal zgomot al imaginii reconstituite de 13 – 19 dB sunt viabile și pot constitui baza unor cercetări în domeniul codării secvențelor video la rate foarte scăzute. Una din principalele direcții viitoare o constituie posibilă combinație a metodei cu tehnicile multirezoluție pentru segmentare, în vederea eficientizării calculului și a stabilității etapei de segmentare și atingerea unor rapoarte de compresie mai mari de 200:1.

Bibliografie

- [AH95] **K. Aizawa and T.S. Huang**, *Model Based Image Coding: Advanced Video Coding Techniques for Very Low Bit Rate Applications*, Proceedings IEEE, vol.83, no.2, p.259-271, Febr. 1995
- [Ale98] **Fl. Alexa**, *Motion Estimation Based on Region*, Proceedings of the Symposium on Electronics and Telecommunications “ETC.’98”, p.199-203, Timișoara, Septembrie 1998

- [CP98] **P. Correira and F. Pereira**, *The Role of Analysis in Content Based Video Coding and Indexing*, Signal Processing, vol.66, no.2, p.125-142, April 1998
- [DM95] **F. Dufaux and F. Mosheini**, *Motion Estimation Techniques for Digital TV: A Review and a New Contribution*, Proceedings of IEEE, vol.83, no.6, p.858-876, June 1995
- [Gir94] **B. Girod**, *Rate Constrained Motion Estimation*, in SPIE Proceedings Visual Communication and Image Processing '94, vol 2308, p.1026-1033, Chicago, September 1994
- [KC98] **O.J. Kwan and R. Chelleppa**, *Region Adaptive Subband Image Coding*, IEEE Trans. on Image Processing, vol.7, no.5, p.632-648, May 1998
- [MCR02] **H. Moon, R. Chellappa, and A. Rosenfeld**, *Performance analysis of a simple vehicle detection algorithm*, Image Vis. Comput., vol. 20, p. 1–13, 2002
- [MDN93] **F. Mosheini, F. Dufaux and H. Nicolas**, *Entropy Criterion for Optimal Bit Allocation Between Motion and Prediction Error Information*, in SPIE Proceedings Visual Communication and Image Processing '93, vol 2094, p.235-242, Cambridge, November 1993
- [MPG85] **H.G. Musmann, P. Pirsch and H.J. Grallert**, *Advances in Picture coding*, Proceeding of the IEEE, vol.73, no.4, p.523-548, Apr. 1985.
- [NM02] **Y. Nie, K.K. Ma**, *Adaptive Rood Pattern Search for Fast Block-Matching Motion Estimation*, IEEE Transactions on Image Processing, Vol. 11, No. 12, December 2002
- [XPC99] **J.-B. Xu, L.-M. Po, and C.-K. Cheng**, *Adaptive motion tracking block matching algorithms for video coding*, IEEE Trans. Circuits Syst. Video.Technol., vol. 97, pp. 1025–1029, Oct. 1999.